PERSPECTIVE
https://doi.org/10.1038/s42256-019-0020-9
nature machine intelligence

Hopes and fears for intelligent machines in fiction and reality

Stephen Cave * and Kanta Dihal *

This paper categorizes some of the fundamental hopes and fears expressed in imaginings of artificial intelligence (AI), based on a survey of 300 fictional and non-fictional works. The categories are structured into four dichotomies, each comprising a hope and a parallel fear, mediated by the notion of control. These are: the hope for much longer lives ('immortality') and the fear of losing one's identity ('inhumanity'); the hope for a life free of work ('ease'), and the fear of becoming redundant ('obsolescence'); the hope that AI can fulfil one's desires ('gratification'), alongside the fear that humans will become redundant to each other ('alienation'); and the hope that AI offers power over others ('dominance'), with the fear that it will turn against us ('uprising'). This Perspective further argues that these perceptions of AI's possibilities, which may be quite detached from the reality of the technology, can influence how it is developed, deployed and regulated.

In the anglophone West, the prospect of intelligent machines is often portrayed in tones of great optimism or equally great pessimism. Regardless of how accurate they are, these portrayals matter, as they create a backdrop of assumptions and expectations against which AI is interpreted and assessed.

There are at least three ways in which these narratives could shape the technology and its impacts. First, they could influence the goals of AI developers. Recently, Dillon and Schaffer-Goddard (manuscript in preparation) have explored this systematically with regard to AI researchers' leisure reading, noting that narratives can "inform and develop research already underway and open up new directions of exploration." Second, narratives could influence public acceptance and uptake of AI systems: for example, a UK parliamentary report[1] notes that those they consulted "wanted a more positive take on AI and its benefits to be conveyed to the public, and feared that developments in AI might be threatened with the kind of public hostility directed towards genetically modified (GM) crops". Third, narratives could influence how AI systems are regulated, as they shape the views of both policymakers and their constituents[2–4].

Given these lines of influence, it is important that narratives about intelligent machines should broadly reflect the actual state and possibilities of the technology. However, the aforementioned parliamentary report emphasized that currently "many of the hopes and the fears presently associated with AI are out of kilter with reality." To understand why this is so, we must first clearly identify and describe those hopes and fears, and second understand why they are prevalent and perpetuated.

This Perspective focusses on the former, with some moves towards the latter. We offer a categorization of what we consider to be the most prevalent hopes and fears for AI, and the dynamics between them. Based on a survey of fictional and non-fictional narratives, we argue that these responses can be structured into four dichotomies, each comprising a hope and a parallel fear. We hope further studies will build on this to examine how and why these narratives are "out of kilter with reality", and the nature of their influence.

Methodology

We set out to categorize strongly prevalent hopes and fears for AI, as expressed in a corpus of popular works, both fiction and speculative non-fiction. We directly examined over 300 works from the twentieth and twenty-first centuries (see Supplementary Information). We also tested whether these categories applied to historical imaginings of intelligent machines as they are described in secondary sources, such as Truitt's *Medieval Robots* and Kang's *Sublime Dreams of Living Machines*[5–9].

Our corpus is not definitive: the range of works engaging with the possibility of intelligent machines is vast and continually growing. But we believe it is large enough to extract key themes. Our primary sources are anglophone Western narratives, plus those narratives that were not originally written in English but are widely available in translation (such as Čapek's *R.U.R.*[10]). We have relied on a variety of indicators in compiling the corpus: for example, for film, we included the IMDB top-35 best-grossing robot films (https://www.imdb.com/list/ls025545074/); for fiction, we have considered older works that are still in print or otherwise deemed classics (such as those that appear in the SF Masterworks series), or more recent works that have won major prizes (such as Leckie's *Ancillary Justice*); for non-fiction, we have looked to relevant works that have attained bestseller status (such as Bostrom's *Superintelligence*, or Kurzweil's *The Age of Spiritual Machines*). We have also included works used as reference points by the media or reports on AI, and those mentioned by the public in a recent survey[11]. Inevitably there will be an element of subjectivity in our selection, and we welcome further suggestions to consider.

In discussing narratives around 'AI', we are conscious that this term was coined only in 1955[12]. Relevant stories both before then and since use a range of other terms. We have therefore not limited ourselves to portrayals that explicitly describe 'AI', but include those that feature machines to which intelligence is ascribed (sometimes, it turns out in the story, falsely). In understanding the term 'intelligence', we follow Margaret Boden's suggestion that it describes "the sorts of things that minds can do"; in particular, she adds, the application of those psychological skills that are used by animals for goal attainment[13]. Under the categories 'artificial' and 'machine' we have similarly cast the net widely, including anything that is built, not born. Common cognate terms describing entities that fall into our corpus are 'robot', 'android', 'automaton' and 'cyborg' (though not all uses of these terms describe relevant entities).

Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, UK. *e-mail: sjc53@cam.ac.uk; ksd38@cam.ac.uk

In distilling our categorizations for the most prevalent hopes and fears for AI, we have considered various previous attempts. A number of authors have written about fears of intelligent machines: Minsoo Kang uses Ann Radcliffe's distinction between terror and horror[14] to categorize negative responses to automata[8]; Kevin LaGrandeur draws from his sources the theme of rebellion[7]; and Daniel Dinello suggests the dystopian themes found in science fiction are a critique of the utopian visions of the technologists themselves[15]. In their analysis of references to AI in the *New York Times*, Fast and Horvitz list both hopes and fears. They include as hopes: improvements to work, education, transportation, healthcare, decision-making and entertainment, as well as a beneficial singularity event and beneficial merging of humans and AI; and as fears: the loss of control of powerful AI, negative impact on work, military applications, a lack of ethics in AI, a lack of progress in AI, a harmful singularity event, and harmful merging of humans and AI[2]. This is a useful survey, although it does not attempt to discern any underlying system to these responses.

We have attempted to distil the positive and negative projections of AI found in our corpus in a way that highlights what we consider to be the most basic themes. There are of course many alternative ways of categorizing these narratives. But we hope this one is at a high enough level to capture the majority of narratives in just a few categories, while still offering some new insight into their underlying structure.

The four dichotomies

We argue that the affective responses to AI explored in the corpus can be placed within a framework of four dichotomies—that is, four hopes and four parallel fears (pictured). We refer to the four hopes as immortality, ease, gratification and dominance. Each is associated with a range of narratives in which intelligent machines have a transformatively positive impact on the lives of some or all humans. Immortality refers to how AI might be used to radically extend life: we give it primacy as staying alive is the precondition for the pursuit of almost any other goal or wish. Once people have that time, ease refers to how AI might grant them the ability to spend it as they wish by freeing them from work. Gratification refers to how AI can help people use that free time, assisting in whatever constitutes pleasurable activity. Finally, AI technologies can be used for what we call dominance, or power over others, as the means to protect this paradisiacal existence.

In claiming that each of these hopes forms one part of a dichotomy, we argue that the utopian visions that they reflect contain inherent instabilities. The conditions required to fulfil each hope also make a dystopian future possible. Thus, the hope for immortality contains the threat of inhumanity—that is, in the pursuit of an ever longer lifespan, a person risks losing their humanity or identity. Ease threatens to become obsolescence, as the desire to be free from work becomes the fear of being put out of work, replaced by a machine. Gratification carries the risk of alienation when in their desire for (artificially) perfect interactions, humans become alienated from each other. And the pursuit of dominance evokes fears of an uprising, as a people's own AI-enabled power is turned—or turns—on them. The factor of control, we argue, balances the hopes and fears: the extent to which the relevant humans believe they are in control of the AI determines whether they consider the future prospect utopian or dystopian.

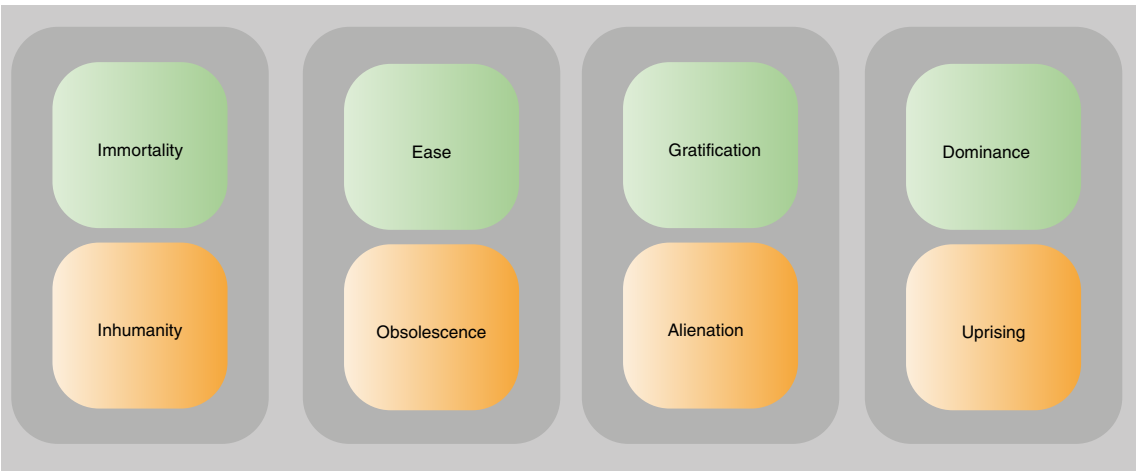
This structure contributes to explaining why responses to AI in the anglophone West are so extreme. The hopeful narratives show the extent to which AI is perceived to be a master tool that can solve problems that have preoccupied humanity throughout history. It represents the apotheosis of the technological dream that humans can use machines to create a paradise on earth[16]. But at the same time, as we explore further below, the idea of creating tools with minds of their own contains (in our imaginings) inherent instabilities. Losing control over such agential machines, or the world they create, is the primary source of the exaggerated fears.

We will now briefly describe these dichotomies and their attendant hopes and fears, illustrating them with examples from our primary sources. Although the examples are used here to illustrate individual hopes or fears, many stories actually address a plurality: the film *Big Hero 6*, for example, expresses all four hopes as well as some of the fears[17].

Immortality versus inhumanity

The pursuit of health and longevity is a basic human drive for the simple reason that it is a precondition for almost anything else one might want. Consequently, humans have always used technology to try to extend lives[18–20]—it is therefore no surprise that one of the great hopes for AI is that it will do just this[3,16,21]. In newspapers, industry reports and elsewhere, AI is portrayed as bringing about a revolution in healthcare, offering better diagnoses, personalized medicine, fewer medical errors and so on[22,23]. Taken to its conclusion, this pursuit of healthier, longer lives becomes the pursuit of indefinitely long lives.

Cave notes two main ways in which people have imagined immortality via AI: transformation and transcendence[24]. In the process of transformation, increasingly sophisticated prophylactics, medicines and prostheses are used to make the body immune to ageing and disease. At its extreme, this is a process of cyborgization, in which the body's unreliable organic parts are replaced by more durable machine parts. By contrast, the process of transcendence



The four dichotomies of hopes and associated fears for AI.

involves total departure from the human body and 'uploading' one's mind onto a machine (which must, by definition, be AI).

These possibilities have been extensively explored in the non-fiction works of the influential technologist Ray Kurzweil, director of engineering at Google, who has written about achieving indefinite lifespans through both transformation and transcendence. In *Fantastic Voyage* he describes what he sees as the increasingly radical medical and other technological interventions that could enable our bodies to keep going indefinitely[25]. In *The Age of Spiritual Machines*[26], he focuses on humans "merging with the world of machine intelligence", resulting in a world in which 'the number of software-based humans vastly exceeds those still using native neuron-cell-based computation.'

But this hope for immortality has its flipside, as Bill Joy put it in his influential 2000 essay 'Why the future doesn't need us'[27]: "on this path our humanity may well be lost." The central concern is whether it is possible for an individual to preserve their identity through the radical metamorphosis that is required to turn an ordinary mortal into something immortal. In one form, this loss of humanity can mean something like loss of human values and emotions. In its more literal form, this fear is that the person hoping for immortality does not really survive at all.

For example, in two episodes, 'Be Right Back' and 'White Christmas', the television series *Black Mirror* critiques different notions of AI-powered survival[28,29]. In the former, a physical replica of a deceased loved one, loaded with his digital data, proves to be a disappointing substitute for the original. In the latter, the character Greta has a digital clone made of herself. The clone is portrayed as having the full personality of the original, but otherwise a much-reduced virtual life. She is informed that her job is to be a digital assistant to the original (who is portrayed as a wholly separate entity—for example, in one scene, asleep while the clone is awake). The narrative makes clear that what the techno-optimists call 'mind-uploading' can also be seen merely as an act of creating copies. Such copies might—in a reduced digital form—outlive us, but they will not literally be us, and so do not offer true survival.

Ease versus obsolescence

Being relieved from the burdens of work is the most ancient hope for intelligent machines. It can be found in the *Iliad*, written around 800 BC, in the form of the golden handmaidens that assist the god Hephaestus[30]. The robot that does our bidding without the complex social and psychological complexities of human servants has been a recurring theme since, both in science fiction and sober predictions of the future. In his essay mentioned above, Joy gives this hope primacy: "the dream of robotics is, first, that intelligent machines can do our work for us, allowing us lives of leisure, restoring us to Eden." Indeed, according to a recent survey exploring public awareness of these dichotomies, this promise that AI will bring a life of luxury and ease is the best known of these hopes for AI[11].

The artificial servant Robby the Robot was the most famous robot of the twentieth century until he was replaced in prominence by the Terminator[31]. In *Forbidden Planet*, Robby is the perfect servant, created by the scientist Morbius as "simply a tool", programmed with "absolute selfless obedience"[32]. Robby is constrained by what Morbius calls 'his basic inhibitions against harming rational beings': he cannot harm a human being, and will overheat and crash when commanded to do so[32]. He is intelligent, but has no will of his own: he will obey humans at all times, with no judgement, being equally eager to protect his master from intruders as to mass-produce bourbon for the spaceship's cook.

But at the same time as people dream of being free from work, they can be terrified of being put out of work. There seems to be a limit to how much leisure time people can tolerate before the fear of becoming entirely obsolete sets in. Work provides people not only with an income, but also with a role in society, status and standing, pride and purpose.

The fear of obsolescence can be divided into two underlying processes. On the one hand, there is the fear of involuntary obsolescence. Jack Williamson's 1947 science fiction novelette 'With Folded Hands' describes a world in which robots protect humans so well, taking away so many jobs that that they consider dangerous or strenuous, that there is nothing left for the human protagonists to do but sit 'with folded hands'[33]. On the other hand, there is the fear of voluntary obsolescence and the long-term effects that that may have on humanity. In *WALL·E*, the human characters seem to be content with their AI-controlled lives, which they spend in immobilizing obesity, in a floating chair, watching screens[34]. For the viewers, this limited life is a dystopian prospect, and they are instead encouraged to identify with the intelligent robots WALL·E and EVE.

Notoriously, a dystopian vision of a future lacking meaningful work was one of the motivations for the 'Unabomber' Ted Kaczynski's violent terrorist campaign against technologists[35].

Gratification versus alienation

Once AI has fulfilled the hopes for longer life and ease, the next goal is to fill all that time with that which brings us pleasure. Just as AI promises to be the perfect servant without the complications of human social hierarchy, so it promises to automate—and thus uncomplicate—the fulfilment of every desire. It could be the perfect companion, for example: always there, always ready to listen, never demanding anything in return. Imaginings of AI are full of such friends: Isaac Asimov's first robot story, 'Robbie' (1939) describes the friendly relationship between a girl and her robot nanny Robbie (not to be confused with Robby from *Forbidden Planet*, although the latter's subservient friendliness was influenced by the former)[36].

In embodied form, AI could also be the perfect lover; fiction tends to present this as a male heterosexual dream. The TV series *Westworld* shows some of the forms such a perfect lover could take: from always-ready prostitutes such as Maeve, to women who have to be courted or subjugated such as Dolores[37]. But even disembodied AI has been portrayed as fulfilling the role of lover: in the film *Her*, Theodore Twombly develops a romantic relationship with his virtual assistant Samantha, who is represented only through her voice[38].

Yet the flipside to the idea of human–AI relationships is that, while some may embrace AI becoming an intimate part of our lives, others may reject the idea of something they perceive to be so unnatural, even monstrous, invading our homes. In robotics, the term 'uncanny valley' describes the revulsion people feel when faced with a replica that is almost human, but not quite[39]. It seems to conjure in us the deep and ancient fear of the doppelgänger, or changeling.

While that fear is based on AI not being human enough, there are also fears around AI being better than humans. If we all have our desires fulfilled by AIs, then we will have become redundant to each other. We might therefore not only become obsolete in the workplace, but even in our own homes and in our own relationships. E. M. Forster anticipated this fear in 'The Machine Stops', in which human interactions are mediated by a machine to such an extent that people never meet each other in person[40]. When the machine stops, they stumble out of their dwellings, disoriented, scared and helpless, and are revolted to come face to face with other humans.

Dominance versus uprising

Finally, the fourth dichotomy concerns power. Once people have long lives and ample free time, and all their desires are fulfilled, they might want to protect this utopia. Indeed, humans have a habit, not just of fighting to protect their favoured way of life, but also of forcing it on others. One major hope for AI is that it can help in retaining or attaining this position of dominance.

Stories of what we now call 'autonomous weapons' are ancient, going back to the bronze giant Talos in the *Argonautica*[41]. In recent times, serious efforts have been underway to make these myths a reality, with significant funding for AI research coming from the military. These themes are also explored in fiction: for example, Iain M. Banks's Culture novels (1987–2012) depict constant clashes between AI-enabled utopians and other life forms resistant to the Culture's imperialism[42]. In other imaginings, AI gives power to the oppressed. In Robert A. Heinlein's *The Moon is a Harsh Mistress* (1966), the inhabitants of the former lunar penal colony make a bid for self-governance aided by a supercomputer that becomes self-aware[43].

The downside of creating autonomous weapons is that such entities might autonomously decide to turn their weapons on their creators. This happens in the very first robot story: Karel Čapek's 1920 play *R.U.R. (Rossum's Universal Robots)*, and has been a persistent theme since[10].

The fears of an uprising are twofold: first, the fear of losing control of AI as a tool—the sorcerer's apprentice scenario. A 2016 White House report[44] highlights the prevalence of this narrative: "In a dystopian vision, super-intelligent machines would exceed the ability of humanity to understand or control. If computers could exert control over many critical systems, the result could be havoc, with humans no longer in control of their destiny at best and extinct at worst. This scenario has long been the subject of science fiction stories, and recent pronouncements from some influential industry leaders have highlighted these fears."

Second, there is the fear that AI systems will turn from mere tools into agents in their own right—what Isaac Asimov called 'the Frankenstein complex'[45]. One of the best-known examples—and beloved of the tabloids—is Arnold Schwarzenegger's T-800 in *The Terminator*[46]. The T-800 is a humanoid robot created by Skynet, an AI that attempts to eliminate humanity as soon as it becomes self-aware: on that day "three billion human lives ended"[47]. The Terminator films symbolize the fears underlying the human hope for dominance by means of AI. Skynet was intended to be an autonomous defence system: it was therefore deliberately given the power and means to destroy other human beings, and at the same time the capacity to develop a will of its own.

As this categorization of the hopes and fears expressed in narratives about AI shows, the idea of creating machines with minds of their own contains inherent instabilities. Of course, there are positive portrayals of intelligent machines, such as the droids R2-D2 and C-3PO in the original Star Wars trilogy. But even here latent dystopian possibilities are visible: the most recent film in that franchise, *Solo: A Star Wars Story*, shows droids standing up for themselves against their human exploiters[48]. The aforementioned survey by Cave et al. shows that public recognition of narratives fits these pairs of hopes and fears: recognition of a positive narrative such as ease equates to recognition of the negative flipside, such as obsolescence[11]. Isaac Asimov was one of the first to recognize this tension, and both critiqued and exploited it in his many robot stories, such as *The Naked Sun*[49]: "One of the reasons the first pioneers left Earth to colonise the rest of the Galaxy was so that they might establish societies in which robots would be allowed to free men of poverty and toil. Even then, there remained a latent suspicion not far below, ready to pop up at any excuse."

Conclusion

In this paper, we have offered a way of approaching the deep-rooted hopes and fears aroused by the prospect of intelligent machines. By structuring them as a series of dichotomies, we hope to have captured both the ambivalence and the strength of feeling (positive and negative) that they invoke. To some researchers in the field, these narratives might seem far removed from the actual power and purpose of the algorithms they are developing. But they nonetheless provide an important context in which their research will be interpreted. As we noted at the start, these narratives around AI stand in a complex causal relationship with the technology itself, both at times inspiring it, and at times trying to reflect it. Yet it is a relationship that also frequently breaks down, in ways that can affect how AI systems will be deployed, adopted and regulated. We hope that understanding the structure and appeal of these framings is a step towards fostering a more balanced discussion of AI's potential.

Received: 26 September 2018; Accepted: 11 January 2019;
Published online: 11 February 2019

References

1. Select Committee on Artificial Intelligence *AI in the UK: Ready, Willing and Able?* (House of Lords, 2018).
2. Fast, E. & Horvitz, E. Long-term trends in the public perception of artificial intelligence. Preprint at https://arxiv.org/abs/1609.04904 (2016).
3. Johnson, D. G. & Verdicchio, M. Reframing AI discourse. *Minds Mach.* **27**, 575–590 (2017).
4. Baum, S. Superintelligence skepticism as a political tool. *Information* **9**, 209 (2018).
5. Mayor, A. *Gods and Robots: The Ancient Quest for Artificial Life*. (Princeton Univ. Press, Princeton, 2018).
6. Truitt, E. R. *Medieval Robots: Mechanism, Magic, Nature, and Art*. (Univ. Pennsylvania Press, Philadelphia, 2015).
7. LaGrandeur, K. *Androids and Intelligent Networks in Early Modern Literature and Culture: Artificial Slaves*. (Routledge, New York, 2013).
8. Kang, M. *Sublime Dreams of Living Machines: The Automaton in the European Imagination*. (Harvard Univ. Press, Cambridge, 2011).
9. Wood, G. *Edison's Eve: A Magical History of the Quest for Mechanical Life*. (Anchor Books, New York, 2002).
10. Čapek, K. *R.U.R.* (Aventinum, Prague, 1920).
11. Cave, S., Coughlan, K. & Dihal, K. 'Scary robots': examining public responses to AI. in *Proc. AIES* http://www.aies-conference.com/wp-content/papers/main/AIES-19_paper_200.pdf (2019).
12. McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Mag.* **27**, 12–14 (Winter, 2006).
13. Boden, M. A. *AI: Its Nature and Future* (Oxford Univ. Press, Oxford, 2016).
14. Radcliffe, A. On the supernatural in poetry. *New Mon. Mag.* **7**, 145–152 (1826).
15. Dinello, D. *Technophobia! Science Fiction Visions of Posthuman Technology* (Univ. Texas Press, Austin, 2005).
16. Noble, D. F. *The Religion of Technology: The Divinity of Man and the Spirit of Invention* (Penguin, New York, 1999).
17. Hall, D. & Williams, C. *Big Hero 6* (Disney, 2014).
18. Gruman, G. J. *A History of Ideas About the Prolongation of Life* (Springer, New York, 2003).
19. Haycock, D. B. *Mortal Coil: A Short History of Living Longer* (Yale Univ. Press, New Haven, 2008).
20. Cave, S. *Immortality: The Quest to Live Forever and How it Drives Civilization* (Crown, New York, 2012).
21. Geraci, R. M. *Apocalyptic AI: Visions of Heaven in Robotics, Artificial Intelligence, and Virtual Reality* (Oxford Univ. Press, Oxford, 2010).
22. Wilson, T. *No Longer Science Fiction, AI and Robotics are Transforming Healthcare* (PwC, 2017).
23. Cockerell, J. Scientists use artificial intelligence to predict how cancers evolve and spread. *The Independent* (2018).
24. Cave, S. in *AI Narratives: A History of Imaginative Thinking about Intelligent Machines* (eds Cave, S., Dihal, K. & Dillon, S.) (Oxford Univ. Press, Oxford, 2020).
25. Kurzweil, R. & Grossman, T. *Fantastic Voyage: Live Long Enough to Live Forever* (Rodale, New York, 2004).
26. Kurzweil, R. *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* (Penguin, New York, 2000).
27. Joy, B. Why the future doesn't need us. *Wired* https://www.wired.com/2000/04/joy-2/ (2000).
28. Harris, O. 'Be Right Back'. *Black Mirror* (Channel 4, 2013).
29. Tibbetts, C. 'White Christmas'. *Black Mirror* (Channel 4, 2014).
30. Cave, S. & Dihal, K. Ancient dreams of intelligent machines: 3,000 years of robots. *Nature* **559**, 473–475 (2018).
31. Telotte, J. P. *Robot Ecology and the Science Fiction Film* (Routledge, New York, 2018).
32. Wilcox, F. M. *Forbidden Planet* (Metro-Goldwyn-Mayer, 1956).

33. Williamson, J. With folded hands. *Astounding Sci. Fict.* **39**, 6–45 (1947).
34. Stanton, A. *WALL·E* (Disney, Pixar, 2008).
35. Kaczynski, T. Industrial society and its future. *The Washington Post* https://www.washingtonpost.com/wp-srv/national/longterm/.../manifesto.text.htm (22 September 1995).
36. Asimov, I. in *The Complete Robot* 164–187 (HarperCollins, London, 1982).
37. Nolan, J. & Joy, L. *Westworld* (HBO, 2016).
38. Jonze, S. *Her* (Sony, 2013).
39. Mori, M. the uncanny valley [from the field]. *IEEE Robot. Autom. Mag.* **19**, 98–100 (2012).
40. Forster, E. M. *The Machine Stops* (The Oxford and Cambridge Review, London, 1909).
41. Apollonius of Rhodes Argonautica Book IV (ed. Hunter, R.) (Cambridge Univ. Press, Cambridge, 2015).
42. Banks, I. M. *Consider Phlebas*. (Macmillan, London, 1987).
43. Heinlein, R. A. *The Moon is a Harsh Mistress*. (Hodder & Stoughton, London, 1966).
44. Preparing for the Future of Artificial Intelligence (Executive Office of the President National Science and Technology Council, 2016).
45. Asimov, I. *The Caves of Steel* (HarperCollins, New York, 1954).
46. Cameron, J. *The Terminator* (Orion, 1984).
47. Cameron, J. *Terminator 2: Judgment Day* (TriStar, 1991).
48. Howard, R. Solo: A Star Wars Story (Lucasfilm, 2018).
49. Asimov, I. *The Naked Sun* (Doubleday, New York, 1957).

Acknowledgements
The authors would like to thank S. Dillon, B. Singler and E. R. Truitt for their helpful comments. This work was funded by a Leverhulme Trust Research Centre Grant awarded to the Leverhulme Centre for the Future of Intelligence.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s42256-019-0020-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence should be addressed to S.C. or K.D.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2019